2-2015

Leakage-resilient password entry: Challenges, design, and evaluation

Qiang YAN
Singapore Management University, qiang.yan.2008@smu.edu.sg

Jin HAN
Institute for Infocomm Research

Yingjiu LI
Singapore Management University, yjli@smu.edu.sg

Jianying ZHOU
Institute for Infocomm Research

Robert H. DENG
Singapore Management University, robertdeng@smu.edu.sg

Citation

YAN, Qiang; HAN, Jin; LI, Yingjiu; ZHOU, Jianying; and DENG, Robert H.. Leakage-resilient password entry: Challenges, design, and evaluation. (2015). *Computers and Security*. 48, 196-211.
Available at: https://ink.library.smu.edu.sg/sis_research/2530

Published in Computers & Security, Volume 48, February 2015, Pages 196-211.
https://doi.org/10.1016/j.cose.2014.10.008

Leakage-resilient password entry: Challenges, design, and evaluation

Qiang Yan [a,*], *Jin Han* [b], *Yingjiu Li* [a], *Jianying Zhou* [b], *Robert H. Deng* [a]

[a] *Singapore Management University, Singapore*
[b] *Institute for Infocomm Research, Singapore*

ABSTRACT

Password leakage is one of the most serious threats for password-based user authentication. Although this problem has been extensively investigated over the last two decades, there is still no widely adopted solution. In this paper, we attempt to systematically understand the challenges behind this problem and investigate the feasibility of solving it. Since password leakage usually happens when a password is input during authentication, we focus on designing leakage-resilient password entry (LRPE) schemes in this study. We develop a broad set of design criteria and use them to construct a practical LRPE scheme named CoverPad, which not only improves leakage resilience but also retains most usability benefits of legacy passwords. Its practicability is further verified by an extended user study.

Keywords:
User authentication
Password leakage
Leakage-resilience password entry
Mobile devices
One-time password

1. Introduction

Even after two decades of attempts to replace password with other alternatives, password is still the most pervasive user authentication mechanism nowadays. Password is easy and cheap to create, use and revoke, which makes it dominant over other authentication mechanisms such as biometrics and smartcards (Bonneau et al., 2012). However, password-based authentication has its intrinsic security weaknesses, among which password leakage is a serious security threat (Long and Wiles, 2008). Password leakage can be caused by various attacks including malware, key loggers, hidden cameras, and timing analysis of user interaction. The consequence of password leakage could be catastrophic, as password-based authentication has been widely used for financial services, online social networks, and other valuable services.

It is widely believed that this threat can be effectively mitigated by using *one-time passwords* (OTPs) generated from *tamper-resistant* hardware tokens (RSA, 2011). However, the applicability of this technique is limited due to the considerable costs of manufacturing, distributing, and managing hardware tokens for service providers, and the costs of carrying hardware tokens for users. As a result, most user accounts in the cyberspace are not protected by hardware-based OTPs. Moreover, hardware-based OTP has its own vulnerabilities such as subjecting to theft (Matsumoto, 2002; Bright, 2011). In order to prevent such vulnerabilities, a hardware-based OTP is usually used together with a password, which is still subject to password leakage attacks. Besides the traditional attacks (Long and Wiles, 2008), the emergence of new

* *Corresponding author*. Current address: Google Inc. Brandschenkestrasse 110, 8002 Zurich, Switzerland. Tel.: +41 0446681800.
E-mail addresses: qiang.yan.2008@smu.edu.sg (Q. Yan), lengshan1983@gmail.com (J. Han), yjli@smu.edu.sg (Y. Li), jyzhou@i2r.a-star.edu.sg (J. Zhou), robertdeng@smu.edu.sg (R.H. Deng).

technologies such as Google Glass[1] further enhance an adversary's capability to capture password without being noticed.

Due to the pervasive use of passwords, extensive research efforts have been conducted on how to design leakage resilient password-based user authentication schemes (Hopper and Blum, 2001; Li and Shum, 2005; Weinshall, 2006; Wiedenbeck et al., 2006; Bai et al., 2008; Kumar et al., 2007; Sasamoto et al., 2008; De Luca et al., 2009a, b; Kim et al., 2010; Bianchi et al., 2011b, a). Despite of all these efforts, there is still no practical and widely adopted solution today. A recent study (Yan et al., 2012) provides strong evidence on the limitations of those schemes that only depend on human cognitive capabilities (Hopper and Blum, 2001; Li and Shum, 2005; Weinshall, 2006; Wiedenbeck et al., 2006; Bai et al., 2008) and concludes that it is necessary to incorporate certain secure channel in the design. The secure channel ensures that at least part of the authentication process should be invisible to an adversary so as to prevent password leakage while maintaining acceptable usability in realistic settings. However, the practicability of using a secure channel in password-based authentication has been considered questionable, as Bonneau et al. (2012) concluded that any user authentication scheme is unlikely to gain traction if it cannot retain comparable benefits provided by legacy passwords.

In this paper, we systematically investigate the underlying challenges of preventing password leakage from both security and usability perspectives. Since password leakage usually happens when a password is input during authentication, we focus on the problem of designing *leakage-resilient password entry* (LRPE) schemes in this study. We develop a broad set of design criteria, which cover three indispensable aspects in LRPE design, including the classic aspect — the tradeoffs between security and usability, and two new aspects — *built-in security*, and *universal accessibility*.

These criteria are then used to guide the design of a practical LRPE scheme named CoverPad, which aims to improve *leakage resilience* of password entry while *retaining most benefits* of legacy passwords. Unlike most prior schemes (Hopper and Blum, 2001; Li and Shum, 2005; Weinshall, 2006; Wiedenbeck et al., 2006; Bai et al., 2008; Kumar et al., 2007; Sasamoto et al., 2008; De Luca et al., 2009a, b), CoverPad is designed for increasingly popular mobile devices equipped with touchscreen, where leakage resilience is achieved by utilizing the gesture detection feature of the touchscreen in forming a *cover* for user inputs. This cover is used to safely deliver hidden messages, which break the correlation between the underlying password and the interaction information observable to an adversary. From the other perspective, CoverPad accomplishes the requirement of retaining comparable benefits of legacy passwords by following our design criteria.

Three variants of CoverPad are implemented and further evaluated with an extended user study. This study includes additional test conditions related to *time pressure*, *distraction*, and *mental workload*. These test conditions simulate common situations for a daily-used password entry scheme, which provides more comprehensive assessment on the practicability of CoverPad. Experimental results show the influence of these conditions on user performance as well as the practicability of our proposed scheme.

The rest of this paper[2] is organized as follows: Section 2 examines closely related research on the LRPE problem. Section 3 introduces the definitions and background of the LRPE problem. Section 4 identifies and analyzes the challenges of designing LRPE schemes. To mitigate the security and usability problems associated with these challenges, we develop a broad set of design criteria, which revisits the classic tradeoffs between security and usability, and extends the scope of security and usability to include built-in security and universal accessibility. Section 5 proposes a practical LRPE scheme for mobile devices equipped with touchscreen. The scheme achieves leakage resilience and retains most benefits of legacy passwords. Section 6 and Section 7 further provide security and usability evaluation to measure the practicability of the proposed scheme. Finally, Section 8 summarizes the contributions of this paper.

2. Related work

In this section, we summarize closely related work on achieving leakage resilience of password entry in three different aspects.

Although the problem of achieving leakage resilience of password entry was proposed two decades ago (Matsumoto and Imai, 1991), it is still a challenge to design a practical solution till now. Early work in this direction (Hopper and Blum, 2001; Li and Shum, 2005; Weinshall, 2006; Wiedenbeck et al., 2006; Bai et al., 2008) focused on designing schemes solely rely on the cognitive capability of human beings. Unfortunately, all such schemes with acceptable usability have been broken (Li and Shum, 2005; Weinshall, 2006; Wiedenbeck et al., 2006; Bai et al., 2008). Recent investigations (Coskun and Herley, 2008; Yan et al., 2012) provided strong evidence for the necessity to construct a partial secure channel for hiding certain user interaction during password entry in order to achieve both security and usability. The establishment of such partial secure channel may require the features only available from new user interface technologies. A few schemes (Kumar et al., 2007; Sasamoto et al., 2008; De Luca et al., 2009a, b; Kim et al., 2010; Bianchi et al., 2011b, a) were designed in this strategy. Among them, our scheme was mostly inspired by the concept of physical metaphor introduced in Kim et al. (2010). Our scheme distinguishes itself from prior work in the sense that it not only achieves leakage resilience but also retains most benefits of legacy passwords, while some of prior schemes (Sasamoto et al., 2008; De Luca et al., 2009b) are flawed in terms of security, and the others incur extra usability costs due to various reasons including: 1) using an uncommon device such as gaze tracker (Kumar et al., 2007; De Luca et al., 2009a), haptic motor (Bianchi et al., 2011b), and large pressure-sensitive screen (Kim et al., 2010), 2) requiring an extra accessory device (Bianchi et al., 2011a), and 3) inoperable in a non-stationary environment (Bianchi et al., 2011b).

[1] Google Glass is currently on sale for $1500 (Google, 2014).

[2] A preliminary version of this paper was presented at the 8th ACM Symposium on Information, Computer and Communications Security (Yan et al., 2013).

On the other hand, the procedure of applying random transformations on a fixed password used in our scheme design is a classic idea to prevent password leakage, but it is not easy to be realized in a *human-friendly* manner without new user interface technologies only available on modern computing devices. These new technologies give our scheme advantages when compared to recently patented schemes. Take GridCode (Ginzburg et al., 2010) as an example, which asks users to memorize extra secrets (besides the passwords) in order to perform the transformations specified in its scheme design, while our scheme does not have such requirement. Another advantage of our scheme is that each character of the password uses a different hidden transformation during an authentication attempt, while GridCode uses the same transformation for all the characters in the password. If a hidden transformation in GridCode is disclosed, the entire password will be exposed. However, if a hidden transformation in our scheme is disclosed, only the single character associated with the transformation will be exposed. These two fundamental differences show both security and usability advantages of our scheme.

In terms of design principles, Roth et al. (2004) proposed to use a cognitive trapdoor game to transform the knowledge of the underlying password into obfuscated responses. Li and Shum (2005) later suggested three other principles including time-variant responses, randomness in challenges and responses, and indistinguishability against statistical analysis. Yan et al. (2012) further extended the coverage by including the design principles against brute force attacks, and provided concrete guidelines against generic statistical attacks. Following this work, a more theoretical analysis on counting-based LRPE schemes was further given by Asghar et al. (2013). Our design criteria complement previous work with detailed guidelines from both security and usability perspectives. Our proposed scheme follows all these design principles to avoid corresponding security flaws.

Bonneau et al. (2012) recently proposed a generic framework for evaluating user authentication proposals and emphasized the importance of retaining the benefits of legacy passwords. This framework is used in our study to guide the scheme design in retaining the benefits of legacy passwords. Other research on password-based user authentication can be found in a recent survey paper (Biddle et al., 2012), which summarized the development of new password schemes in the past decade.

3. LRPE problem overview

In this section, we introduce the definitions of the LRPE problem, and then define its threat model and the scope of this paper.

3.1. *Definitions*

In general, an LRPE scheme allows a human *user* to be authenticated to a (local or remote) computer *server* in a secure manner. During registration, user and server agree on a *password*, where each element contained in the password is referred to as a *password element*. A password element can be a text character, an image, or any symbol in a notational scheme. The user later uses her knowledge of the password to generate *responses* to *challenges* issued by the server to prove her identity. This process is referred to as *password entry*. In the case of legacy passwords, the user directly enters her plaintext password so that an adversary may capture the password via various attacks including malware, key loggers, and hidden cameras. *Password leakage* is the threat that a user's password is directly disclosed or indirectly inferred. The purpose of an LRPE scheme is to establish a *leakage-resilient* environment to mitigate or prevent password leakage during password entry.

An authentication scheme is *not* considered as an LRPE scheme if a user only *transcribes* the response generated by a tamper-resistant device (RSA, 2011). Such a scheme addresses a different problem which verifies a user to be the person who possesses the device. It is usually used together with legacy passwords or biometrics so as to mitigate the risk of unauthorized access to the device, and may still be subject to the password leakage threat. A secure LRPE scheme can be used to effectively mitigate this threat by strengthening password entry so as to construct a more secure multi-factor user authentication.

3.2. *Threat model*

Various potential attacks need to be addressed in the design of LRPE schemes. An adversary may use malware, key loggers, or other sophisticated mechanisms to capture messages delivered between user and server and infer the underlying password. Prior proposals on LRPE schemes can be categorized according to whether or not a secure channel is used in the authentication process. There are quite a few LRPE schemes in the literature which are designed solely based on human cognitive capabilities without using any secure channel (Hopper and Blum, 2001; Li and Shum, 2005; Weinshall, 2006; Wiedenbeck et al., 2006; Bai et al., 2008). However, none of those schemes are both secure and usable (Golle and Wagner, 2007; Li et al., 2009; Asghar et al., 2010; Perkovic et al., 2011; Yan et al., 2012). It is shown by Yan et al. (2012) that an LRPE scheme must rely on the existence of certain secure channel to achieve both security and usability.

Although it would be difficult to establish a standard secure channel protecting all messages delivered between user and server, it is possible to construct a partial secure channel. The requirement of a *partial* secure channel is weaker than a standard secure channel, as it only requires that a portion of messages delivered between user and server be invisible to an adversary. The use of a partial secure channel ensures that the leakage resistance of an LRPE scheme is preserved even after allowing an adversary to observe most messages during password entry as long as certain critical messages are not disclosed. A partial secure channel is usually *unidirectional* either from server to user or from user to server.

In the presence of a partial secure channel, it is possible to achieve the optimal security objective, *no password leakage* during password entry. No password leakage with a partial secure channel means that if the portion of messages protected by the partial secure channel is not disclosed, the most efficient attacks to learn the underlying password are online dictionary attacks. This study focuses on LRPE schemes using such a partial secure channel and excludes LRPE schemes without using any form of secure channel unless explicitly mentioned.

In addition to the attacks mentioned above, password leakage may happen due to other types of attacks, such as social engineering, phishing or even non-technical attacks such as dumpster diving (Long and Wiles, 2008). Although their mitigation technologies such as secure URL checker and spam filter have become standard components of modern computer systems, some of these attacks cannot be completely thwarted by technical solutions alone and they are orthogonal to the password entry problem. Another example is the database reading attack, where the adversary intrudes into the back-end databases to compromise all user passwords. These attacks are beyond the scope of this paper.

4. Challenges behind LRPE problem

In this section, we investigate the underlying challenges of designing secure and usable LRPE schemes. Based on this analysis, we develop nine design criteria, which are used later to guide our scheme design. These criteria are organized in three different aspects as introduced in the following subsections.

4.1. *Relations between security strength and usability costs*

The relations between security and usability in LRPE design are usually considered as tradeoffs, as most security enhancements are always associated with certain usability costs. Since extra usability costs seem inevitable, it is not easy to approach our design objective that requires retaining the *comparable* usability benefits of legacy passwords (Bonneau et al., 2012). This difficulty stems from a crucial fact that unassisted human beings are not capable of performing all the operations as required in secure LRPE design (Yan et al., 2012). Among these capabilities, memory capacity and mental calculation are essential. Memory capacity limits the length of passwords or other user secrets that can be memorized by a user, while mental calculation constrains the complexity of arithmetic or logic calculation that can be performed with human brain only. These limitations are fundamental for unassisted human beings (Yan et al., 2012; Asghar et al., 2013), but they can be overcome with the assistance of external trusted components that perform the operations unaffordable to human beings.

The usage of trusted components conceptually establishes a partial secure channel between user and server. The various ways of using these trusted components create various tradeoffs in LRPE design. Our analysis starts from two extreme cases, which assume that a user assisted by trusted components always has sufficiently large memory capacity or sufficiently powerful calculation capability, so that an LRPE scheme can mainly rely on only one of these two capabilities and minimize the requirement on the other.

In the first case, an assisted user can use the extreme memory capacity to store a long sequence of random digits as password and a cursor pointing to the first unused digit in this sequence. Every time, a user uses the next n unused digits from the position pointed by the cursor to pass user authentication, where n is a configurable positive integer. This design imposes the *minimum* requirement on mental calculation and works exactly like hardware OTPs. In the second case, an assisted user can use the extreme calculation capability to perform a chosen plaintext attack (CPA)-secure encryption or secure one-way hash function on a password of minimum length, so that only the *minimum* requirement on memory capacity is imposed. Then the user uses the computed ciphertext as the response for user authentication, where it would be computationally infeasible for an adversary to infer the actual password from the ciphertext. However, if a user needs to enter the actual password for each ciphertext generation, this scheme is still subject to common password leakage attacks during password entry. In order to mitigate this threat, the password has to be pre-stored in a trusted component, which reduce the scheme into the one used in the first case.

It is easy to see that the solutions for both extreme cases can be reduced to hardware-based OTPs, as both of them require tamper-resistant hardware to safely store the plaintext of password and ask a user to transcribe a generated response to prove her identity. The procedure of transcribing the response from a trusted component essentially forms a partial secure channel from server to user that delivers an OTP that is assumed only visible to the legitimate user. Correspondingly, these solutions also have similar usability costs and security vulnerabilities as hardware-based OTPs, where the password is *persistently* stored in a trusted component. Once an adversary gains the access to this trusted component even when a user is *not* using it, these solutions will be compromised. This is an inherent challenge in the LRPE problem. Theoretically, the best security protection against this threat is to ensure that password leakage attacks are feasible *only when* a user is using the password, which introduces the first criterion.

> **C1**: *A secure LRPE scheme should not persistently store any password or other information that can be used to infer a password.*

Since the solutions under two extreme cases analyzed previously do not satisfy C1, we need to search in the remaining design space so as to avoid persistently storing password. It indicates that a secure LRPE scheme will impose more than the minimum requirement on both memory capacity and mental calculation, even with the assistance of trusted components. The actual usability costs depend on how these trusted components are used to construct a partial secure channel and how efficiently a user interacts with this channel and other interaction channels used in the design.

There are two types of interaction channels which can be used in a typical LRPE scheme, input channels and output channels. Existing user interfaces require that a user gets inputs from vision (Kim et al., 2010; Kumar et al., 2007; De Luca et al., 2009a; Sasamoto et al., 2008; De Luca et al., 2009b), acoustics (Bianchi et al., 2011a), or haptics (Bianchi et al., 2011b), and provide outputs via acoustics or motion (Kim et al., 2010; Kumar et al., 2007; De Luca et al., 2009a; Sasamoto et al., 2008; De Luca et al., 2009b; Bianchi et al., 2011a, b). For the input channel, evidences from psychology show that vision is the fastest channel to reliably collect information for non-blind users. This phenomenon is called *visual dominance*. In perception and information processing, vision has been shown to dominate over acoustics (Colavita, 1974) and haptics (Gibson, 1937). For the output channel, motion is shown to be a more reliable and faster channel compared to the acoustics channel as average human beings have better control over body, especially hand, than sound (Stifelman et al., 1993) and it has better resistance against environmental noises. Among all the possible motions, clicking (Kim et al., 2010; Sasamoto et al., 2008; De Luca et al., 2009b; Bianchi et al., 2011a, b) is the simplest which only requires the user to move one finger without a high precision control as required by other motions like shaking in a specific way. Hence, the optimal choice for interaction channels in general is vision for input, and clicking for output. Any other choices for interaction channels may be considered *low efficiency* unless they are designed for specific application scenarios. Therefore, the second criterion is introduced to characterize the efficiency of the chosen interaction channels.

> **C2**: *A usable LRPE scheme should choose the most efficient interaction channels that have 1) high bandwidth for efficient message delivery, and 2) high reliability and minimum demand on human capabilities so that target users can use them easily in practice.*

In realistic settings, an interaction between user and user interface may be captured through multiple *leakage* channels from an adversary's perspective. For example, a clicking action on a keypad may be intercepted from a vision channel, where an adversary installs a hidden camera, or from a haptic channel, where an adversary installs a sensitive haptic board above the original keypad. Fig. 1 characterizes the common leakage channels in a typical attack scenario. The attacks exploiting these channels include *vision*-based eavesdropping, *haptics*-based eavesdropping, and *acoustics*-based eavesdropping, which are described as follows.

For vision-based attacks, an adversary may infer the actual password by observing the movement of fingers even without direct line-of-sight on the screen display. This capability is significantly enhanced with emerging augmented-reality accessory like Google Glass, which is a small wearable glass transferring real-time video captured by a tiny camera to a server and displaying the analyzed results received from the server.

Haptics-based attacks are most likely to happen when users use public computer kiosks equipped with touchscreen. In particular, mobile devices, such as iPad, have been used as public computer kiosks as observed in museums, restaurants, and hotels. This provides an incentive for an adversary to install a physical "touch" logger. Although such touch logger has not been observed in the wild, it is technically feasible to implement as other physical key loggers. Considering that the thickness of touchscreen in Samsung Galaxy S3 is just 1.1 mm (Androidcommunity, 2012), it may not be noticeable to users if an extra physical touch logger is installed, especially when it is installed on a kiosk device inside a thick anti-theft box.

The effectiveness of acoustics-based attacks depends on whether user actions can be distinguished by their tone patterns. For example, different tones are played when a user dials different numbers on an old-style phone (Schenker, 1960). Due to environmental noises, acoustics-based attacks are usually not as effective as vision-based attacks and haptics-based attacks.

All the above attacks need to be properly addressed in a secure LRPE scheme by protecting the corresponding leakage channels, which is also the major security objective in our design. If it is impossible to completely eliminate the threat from a certain leakage channel, this leakage channel should be transformed into a partial secure channel by incorporating trusted components in the design. Every partial secure channel relies on certain trust assumptions to reduce the corresponding threat to a manageable level. Consequently, the attacks become possible only under certain conditions, and these conditions are hard to achieve by an adversary or easy to detect by a user. Although it is difficult to quantify the leakage resistance of different partial secure channels in general, the risk of password leakage can be reduced if a fewer number of partial secure channels are involved in a scheme design. This leads to the third criterion.

> **C3**: *A secure LRPE scheme should minimize the number of partial secure channels used in the design.*

Besides interaction channels, the construction of a partial secure channel also involves non-interactive components including middleware, operating systems, network, and other application logic that may intercept sensitive messages

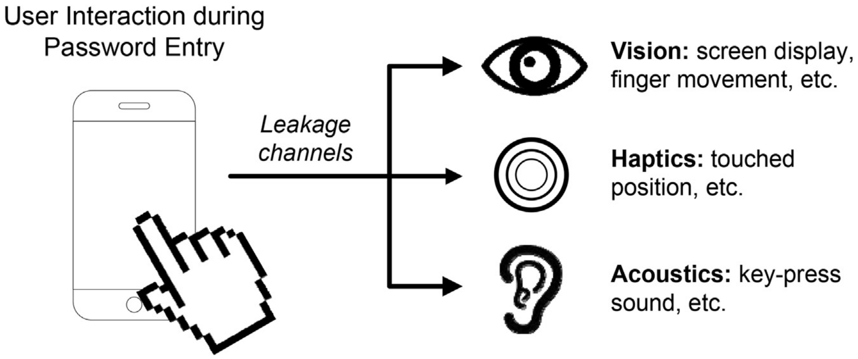

User Interaction during Password Entry
Leakage channels
Vision: screen display, finger movement, etc.
Haptics: touched position, etc.
Acoustics: key-press sound, etc.

Fig. 1 – Leakage channels in a typical attack scenario.

delivered during password entry. An adversary exploiting these components is usually required to access *internal* states such as the memory of the computing device used for user authentication. Typical attacks include *logic key loggers*, *malware*, and *network eavesdropping*, which are common to all password-based user authentication schemes.

Unlike previous *external* eavesdropping attacks exploiting interaction channels, the protection mechanisms against these attacks on non-interactive components do not affect usability and do not introduce new tradeoffs between security and usability. On the other hand, existing solutions such as application sandbox (Microsoft, 2014; Begemann, 2012) are available to effectively defend against these attacks, though it takes time for them to replace legacy vulnerable systems. These solutions are independent on user interaction during password entry so that they can be adapted to any user authentication schemes including LRPE schemes. Therefore, like most prior research (Kumar et al., 2007; Sasamoto et al., 2008; De Luca et al., 2009a, b; Kim et al., 2010; Bianchi et al., 2011b, a), we will not directly address these attacks in our scheme design.

4.2. *Built-in security*

Built-in security requires that the security strength of an LRPE scheme should not rely on optional user behavior. If a scheme requires a user to perform an optional action to achieve its security strength, this security strength is unreliable as the user may not act appropriately. This could be caused by the inconsistency with personal habits and the sensitivity on violations of social norms, which are discussed below.

There are two common inconsistencies between users' habits and security design. The first one is *impatience*, which means a user may not perform any optional actions which she is supposed to perform. Some common optional actions such as reading a manual, and checking the integrity of input device may make users impatient. A typical example is Error-Correcting-Challenge (Hopper and Blum, 2001), which is the only existing scheme that is designed to defend against an *active* adversary. The adversary is allowed to arbitrarily manipulate the environment for password entry, such as modifying a challenge issued by a legitimate server. This scheme requires a user to verify the integrity of a challenge by testing the linearity of the digits shown in the challenge before answering it. A user should answer the challenge only after the challenge passed a sufficient number of linearity tests on hundreds of digits. During this process, a user may become impatient due to the high demand on mental calculation. As a result, its security strength may not be achieved. To avoid this kind of security threat, the fourth criterion is introduced.

> **C4**: *A secure LRPE scheme should not rely on any optional user actions to achieve its security objective.*

The second inconsistency is about users' inability of *generating random numbers*, where certain LRPE schemes rely on users to make random choices. For example, the LPN scheme (Hopper and Blum, 2001) asks a user to calculate the responses using two different algorithms randomly. Given a challenge, with a fixed probability η, the user chooses the first algorithm; otherwise she uses the second algorithm. The user passes the authentication if the ratio of correct responses generated by the first algorithms is close to η. The leakage resistance of this scheme relies on the randomness in users' choices between the two algorithms, which may be significantly undermined if a user always follows a fixed pattern to choose between these two algorithms. It is usually difficult for average users to make such "random" choices (Aumann, 1974) specified by this scheme design. This counterexample shows that if the security strength of an LRPE scheme relies on an expected pattern of user behavior, the scheme design should ensure that a user can always follow the expected pattern; otherwise, the scheme is vulnerable. This leads to the fifth criterion.

> **C5**: *A secure LRPE scheme requiring a user to perform certain actions in a specific behavior pattern should ensure that a user can pass the authentication only when the expected behavior pattern is detected.*

Besides personal habits, social norms are also common concerns impeding a user from performing certain optional protection actions. A recent field study on ATM usage (De Luca et al., 2010) found that a user is not willing to shield a keypad if she is accompanied by her friends. The user may think that a shielding gesture would be misinterpreted as a sign of distrust to her friends. This situation is more likely to happen among users who have an intimate relation with each other. Social norms may vary with different cultures, but their impact on LRPE schemes is similar, which may prevent users from performing certain optional protection actions required by LRPE schemes. Hence, a secure LRPE scheme should follow the previous criteria (C4 and C5) to make necessary actions mandatory so as to achieve security objectives. This is also a solution to avoid potential misinterpretation on social norms.

User education may alleviate the problems related to built-in security to some extent, but the outcome is uncertain. A user may still make mistakes or be overconfident. Any LRPE scheme should not rely mainly on user education to achieve its security. It is necessary to convert optional actions into compulsory actions if they are critical to the security strength of LRPE schemes.

4.3. *Universal accessibility*

Universal accessibility is intended to benefit the majority of users in the design of LRPE schemes. Specifically, it requires a scheme to be accessible even in a non-ideal environment such as situations when a user is not able to use all her capabilities or when environmental noise is high. Traditional laboratory user study in ideal environment for unhampered users may not be sufficient to fully evaluate the usability of LRPE schemes in practice. We discuss three general aspects of universal accessibility below.

Beneficiary scope is the first aspect that specifies who has the capabilities to use an LRPE scheme. The success of legacy passwords is largely attributed to its wide beneficiary scope, as it imposes minimum requirement on human capabilities in a general sense. Anyone who can see and move a single finger can use legacy passwords. A narrower beneficiary scope means some current users of legacy passwords cannot use the

LRPE scheme. A practical LRPE scheme should preserve a similar beneficiary scope. Any LRPE scheme that requires extra human capabilities may not be appealing to the majority of users. For example, the PressureGrid scheme (Kim et al., 2010) requires precise cooperation of multiple fingers. This cooperation would be difficult especially for elders, children, and those with physical (not *cognitive*) disability such as a person who loses one of her fingers. So the sixth criterion is introduced to address this aspect.

> C6: *A usable LRPE scheme should minimize the requirement on human capabilities.*

The second aspect is device availability. Any LRPE scheme runs with at least one device, where the user uses a system protected by the LRPE scheme. This device is referred to as the *primary device*. Some existing LRPE schemes (Bianchi et al., 2011a; De Luca et al., 2009b) also require an extra device to form a partial secure channel, which is referred to as the *secondary device*. The use of secondary device lowers device availability, even if the device is free of charge. This is because the secondary device must be carried by users and it is subject to extra risks such as theft, which in turn may cause security or accessibility problems. This requirement on device availability introduces the seventh criterion.

> C7: *A usable LRPE scheme should avoid the use of secondary device and focus on reusing the existing features of the primary device.*

Even if a primary device is equipped with sufficient features to support an LRPE scheme, it usually has its own functional limitations. One of common limitations is the screen size. As mobile devices such as smartphones and tablets are becoming the major computing devices used by ordinary users, it is necessary for a usable LRPE scheme to support these devices that are usually equipped with a small screen. Visual redundancy shown in the previous schemes (Wiedenbeck et al., 2006; Bai et al., 2008) should be avoided, which is further addressed in the eighth criterion.

> C8: *A usable LRPE scheme should control the number of visual elements that are displayed simultaneously on the screen for a better adaptation to various computing devices.*

The last aspect is environmental adaptation. Laboratory user study is usually conducted in a quiet room and each user is given sufficient time to perform a single task in each test. However, this may not be the case in daily usage. Users may act differently when they do not have peace in mind or not stay in a quiet room. Below we summarize common environmental factors which affect users' perception of security. 1) *Impact of time pressure*: a user tends to act hastily under time pressure, which may lead to mistakes. 2) *Impact of distraction*: unexpected distraction interferes with a user's mind when answering challenges. 3) *Impact of mental workload*: mental workload consistently interferes with a user's mind during answering challenges. 4) *Impact of environmental noise*: environmental noise may render certain interaction channels such as acoustics and haptics imprecise or even unusable (Bianchi et al., 2011b). 5) *Impact of hampered capability*: a user's capability may be hampered even if she is not handicapped. For example, a user may only use one hand in authentication when she uses the other hand to carry a bag. These environmental metrics are important in the evaluation of LRPE schemes so as to obtain credible results in real-world scenarios, which also introduce the last criterion.

> C9: *A usable LRPE scheme should be operable in a non-ideal environment.*

5. CoverPad design

5.1. *Conceptual design*

Guided by the design criteria developed in the previous section, we present the design of CoverPad, which is designed to improve leakage resilience of password entry while retaining most benefits of legacy passwords. CoverPad leverages on the touchscreen feature of mobile devices. Its conceptual design is

Setup:
A server and a user agree on a k-length password $pwd = (a_1, a_2, \ldots, a_k)$, where a ***password element*** *$a_i = pwd[i]$ belongs to an alphabet with size w. It is allowed that $a_i = a_j$, for $i \neq j$.*

Password Entry:
For each i from $[1, k]$:
Step 1: *The touchscreen shows a keypad with all the elements in the alphabet.*
Step 2: *The user is asked to perform a hand-shielding gesture to read the hidden transformation $T_i(\cdot)$ protected by the hand-shielding gesture. $T_i(\cdot)$ will immediately disappear if the gesture is no longer detected.*
Step 3: *The user clicks on response element e_i, where $e_i = T_i(a_i) = (a_i + r_i \mod w)$, where r_i is a random number drawn from a uniform distribution. A new random number r_i is generated for each round i. The hand-shielding gesture is not required for this step.*

Fig. 2 – Conceptual design of CoverPad.

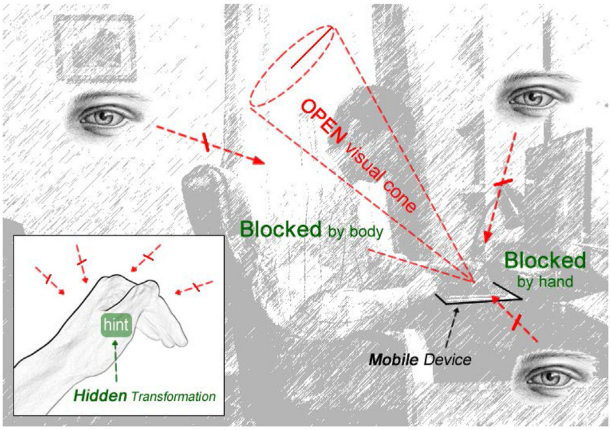


Fig. 3 – The hand-shielding gesture and its effectiveness.

shown in Fig. 2, where a hidden transformation $T_i(\cdot)$ is a random mapping $\Omega \rightarrow \Omega$, where Ω is the set of all individual elements contained in the password alphabet.

An example of using CoverPad is given as follows. Suppose a user has a k-length password. At the beginning of password entry, the user performs a hand-shielding gesture to view the current hidden transformation T_1 for the first character a_1 in her password. Then, she applies T_1 to a_1 and enters the transformed response e_1. This procedure repeats for each password element a_i so that the overall computational complexity for a user will be $O(k)$ for k transformations applied in total. During the whole password entry, T_i disappears immediately once the gesture is not being detected. A user can always view T_i by performing the gesture again before inputting e_i.

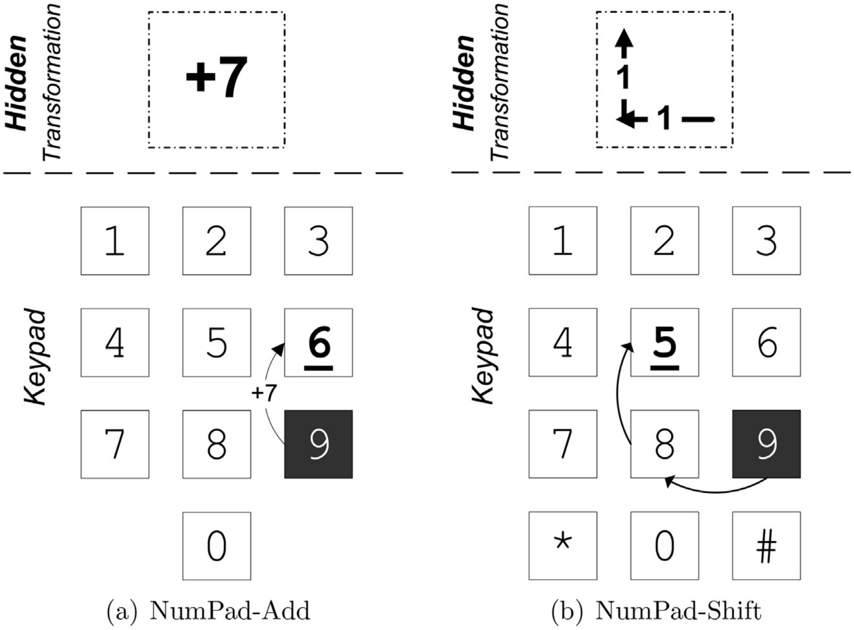
Fig. 3 shows how to correctly perform a hand-shielding gesture. This gesture restricts the vision channel to a small visual cone. This visual cone is not accessible to an adversary unless the adversary's eyes are close enough to the user's head, which makes the adversary easily exposed. A hidden camera near the line of sight may help capture the hidden transformation. However, it needs to be adjusted according to the user's height and current position, which may lead to user's awareness. On the other hand, the observable responses for the same password element are uniformly randomized. Thus, CoverPad is also immune to haptics-based eavesdropping. As long as the hidden transformation is not revealed together with the corresponding response, observed interaction provides no valuable information for an adversary to infer the actual password. A proof about this security property will be given in Section 6.

CoverPad follows all our design criteria proposed in the previous section. It does not store the password persistently (C1); it uses vision for input and clicking for output, which is the optimal choice for interaction channels suggested by C2. The only partial secure channel involved is vision (C3). For built-in security, it does not require any optional user actions (C4) and it works only when the expected hand-shielding gesture is correctly detected (C5). For universal accessibility, the human capabilities required by CoverPad are exactly the same as legacy passwords (C6) and no extra device is needed (C7). Furthermore, the requirements on screen size and operating environment are also the same as legacy passwords (C8 and C9).

5.2. *Implementation variants*

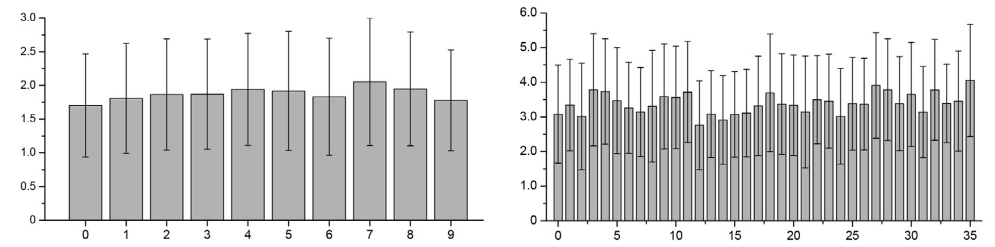
We provide three variants of CoverPad, *NumPad-Add*, *NumPad-Shift*, and *LetterPad-Shift*. They implement different features tailored for users with various skill sets. Fig. 4 illustrates these variants.

In NumPad-Add, the alphabet of password consists of digits 0 to 9 only. The hidden transformation is performed by



(a) NumPad-Add (b) NumPad-Shift

Fig. 4 – Demonstration of CoverPad variants. Note that LetterPad-Shift is not shown here, which is similar to NumPad-Shift.



(a) Timing deviation for NumPad-Add (b) Timing deviation for LetterPad-Shift

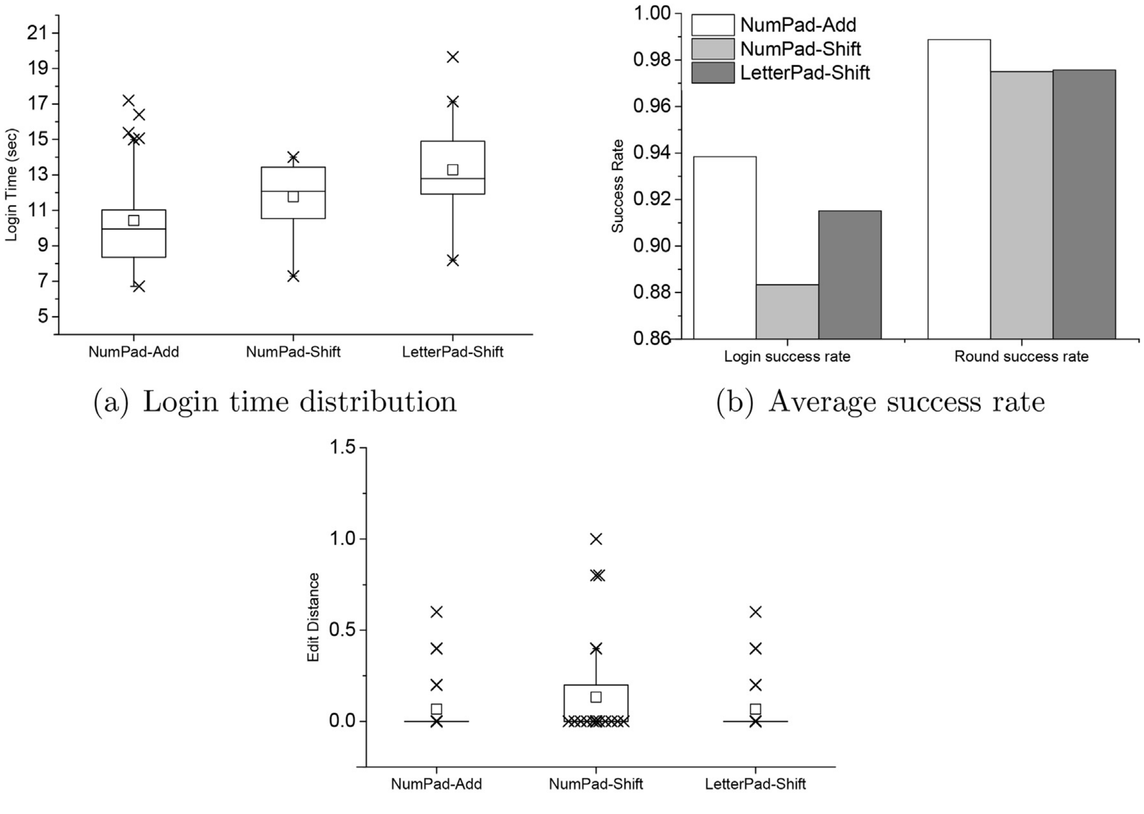
Fig. 5 – Timing deviations and distributions for entering each password element. The results of NumPad-Shift are similar to the results of NumPad-Add shown in these figures.

adding a random digit to the current password element and then mod10 if the sum is larger than 9, where the value of the random digit ranges from 0 to 9. For example, the correct response for the first round is $6 = (9 + 7)\mathrm{mod}10$ given password 934567 and the hidden message '*plus 7*'.

In NumPad-Shift, the alphabet of password consists of digits 0 to 9 only. The hidden transformation is performed by *shifting* the location of the current password element by X-offset and Y-offset, where the offset values are randomly taken from $\{-1,0,1\}$ for X-offset, and $\{-1,0,1,2\}$ for Y-offset. For a 3×4 keypad design shown in Fig. 4(b), the transformed response for a_i is calculated as $pad[x(a_i) + \Delta x mod3][y(a_i) + \Delta y mod4]$, where Δx is the X-offset, Δy is the Y-offset, and $x(a_i)$ is the X-index of a_i, and $y(a_i)$ is the Y-index of a_i. For example, the correct response for the first round is 5 if the password is 934567 and the hidden message is '*move left by 1 step and move up by 1 step*'.

Note that two extra keys * and # are added to the keypad; otherwise, the distribution of hidden transformations is not uniform on the keypad layout. The proof for the necessity of these two keys is given in Yan et al. (2013).

In LetterPad-Shift, the alphabet of password consists of letters *a* to z and digits 0 to 9 (36 elements in total). The hidden transformation is the same as NumPad-Shift. The offset values are randomly taken from $\{-2,-1,0,1,2,3\}$ for both X-offset and Y-offset for a 6×6 keypad design. The transformed response for a_i is calculated as $pad[x(a_i) + \Delta x mod6][y(a_i) + \Delta y mod6]$ in a similar way as for NumPad-Shift.



(a) Login time distribution

(b) Average success rate

(c) Edit distance distribution

Fig. 6 – Average login time, success rate, and edit distance under the normal condition.

6. Security analysis

CoverPad inherits all the existing security benefits of legacy passwords, such as no password stored on a local device. Besides those benefits, it extends the security strength against password leakage, particularly against common external eavesdropping attacks and side-channel attacks, as analyzed in this section.

6.1. *External eavesdropping attacks*

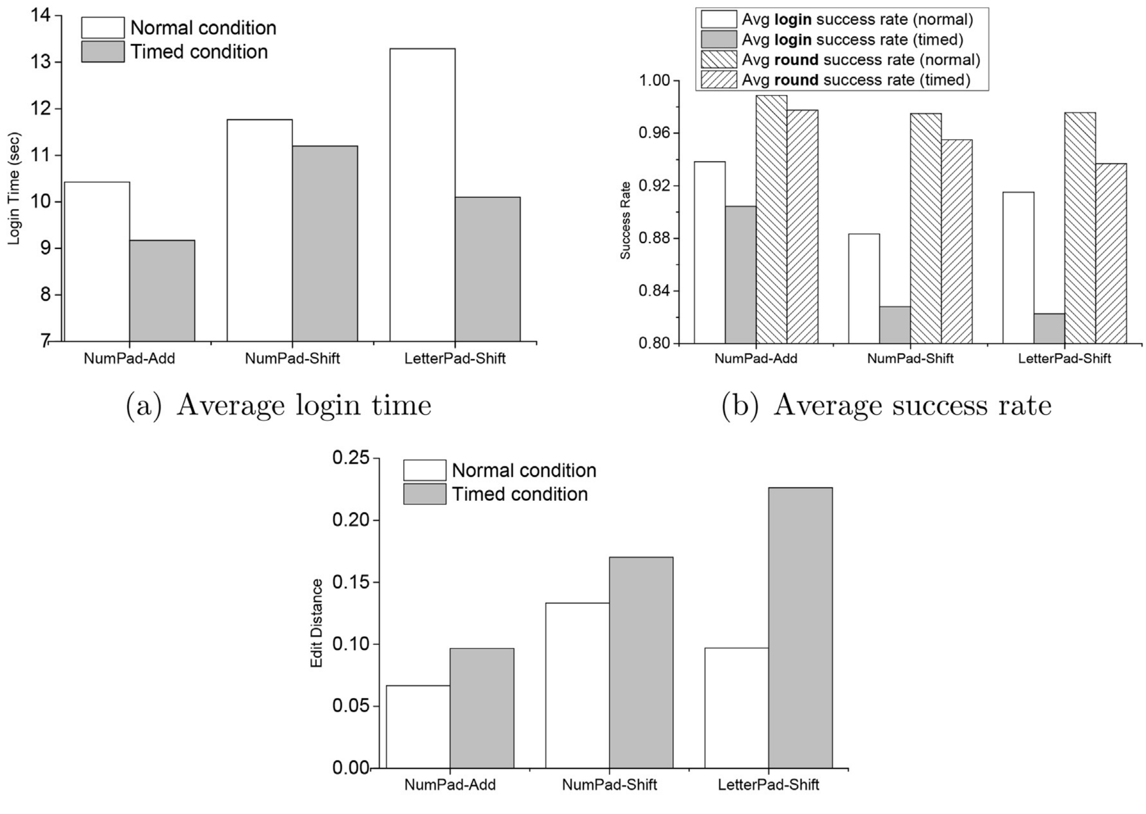
Common external eavesdropping attacks leading to password leakage may exploit vision, haptics, or acoustics channel as analyzed in Section 4.1. For CoverPad, an adversary using these attacks can observe *at most* a complete response key sequence pressed by a user, while the hidden transformation is protected by our design. From this key sequence, the adversary knows the i-th pressed key is decided by the i-th element in the password. However, the adversary cannot further infer what the i-th password element is, as proved as follows.

Proof. Given a pressed key e_i, and two password elements a_x and a_y in a w-sized password alphabet, let $Pr(e_i|a_x)$ and $Pr(e_i|a_y)$ be the probabilities for e_i being pressed when the underlying password element are a_x and a_y, respectively. We have $Pr(e_i|a_x) = Pr(e_i = a_x + r_i \mathrm{mod} w) = Pr(r_i = e_i - a_x \mathrm{mod} w) = Pr(r_i = C \mathrm{mod} w) = 1/w = Pr(e_i|a_y)$ for any i, x, and y, where C is a constant integer randomly drawn from a uniform distribution. Therefore, a sequence of pressed keys observed by an adversary is equivalent to a random sequence, which is similar to a ciphertext generated by a one-time pad.

Using a partial secure channel where the hidden transformation is protected by the hand-shielding gesture, our scheme achieves no password leakage. As long as the hidden transformation is not disclosed together with the corresponding response, an adversary cannot infer any information about the underlying password (except password length) even after an infinite number of observations.

6.2. *Side-channel attacks*

In reality, it is possible for an adversary to exploit subtle side-channels to collect password information during password entry. These attacks are not usually considered in common threat models (Hopper and Blum, 2001; Li and Shum, 2005; Weinshall, 2006; Wiedenbeck et al., 2006; Bai et al., 2008; Kumar et al., 2007; Sasamoto et al., 2008; De Luca et al., 2009a, b; Kim et al., 2010; Bianchi et al., 2011b, a). A typical side-channel attack is timing analysis (Song et al., 2001), which analyzes the patterns in the response time of entering individual password elements. The preliminary results of our scheme against timing analysis are given in Fig. 5. For the timing deviation shown in Fig. 5(a) and (b), each bar with x-value i represents the average response time for entering the transformed responses for a specific password element i. These



(a) Average login time

(b) Average success rate

(c) Average edit distance

Fig. 7 – Impact of time pressure.

results show the range and the distribution of the response time for entering different password elements are almost overlapped. Further analysis shows that the results of timing variance for individual participants are similar to the results shown in the above figures. This indicates that timing analysis is not a major concern for our scheme, though it is difficult to completely prevent such attacks due to inevitable human behavior patterns during password entry. Detailed analysis on side channel attacks is out of the scope of this paper.

7. Usability evaluation

7.1. *Methodology*

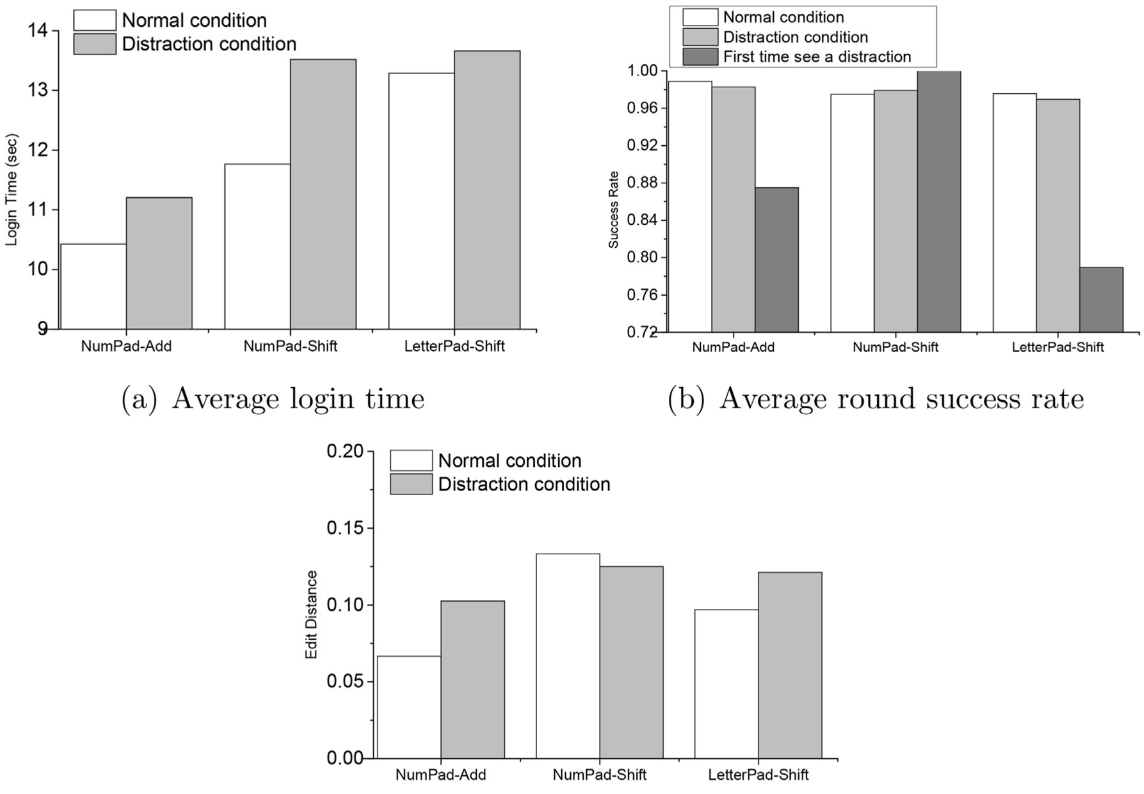
The participants in our user study are recruited from undergraduate students in our university. There are 61 participants in total, 30 male and 31 female, with age range between 20 and 25. These participants come from five different departments, in which 42 of them have a social science or business related background, and the remaining 19 have a computer science or information technology related background. Each participant is paid with 10 dollars as compensation for their time.

The user study is conducted in a quiet room. The experiments use a within-subjects design. Each participant is asked to use all three variants as three *test groups*. These variants are implemented on Apple iPad, which are referred to as *schemes* in this section. The order of the schemes is randomized to avoid the learning effect that affects the performance for a specific scheme. For each test group, a user is required to memorize a *randomly generated* password in the beginning. The password strength is set to be equivalent to 6-digit PIN, which is the strength that is generally used by the OTPs for online banking and corporate login. Correspondingly, the password length is four for LetterPad-Shift, and six for both NumPad-Shift and NumPad-Add. The participants learn how to use a scheme by an interactive step-by-step tutorial.

In each test group, there are six tests simulating additional *test conditions* that evaluate the influence of time pressure, distraction, and mental workload. The details of these test conditions are described in the next subsection. The order of these tests is also randomized in order to avoid the learning effect. All three test groups consist of 18 tests in total. At the end of the user study, the participants are given a questionnaire using 5-point Likert scale to collect their perception on the schemes. The whole user study takes 35~50 minutes to complete. More details about the experiment process are provided in Yan et al. (2013).

7.2. *Simulating various test conditions*

In order to simulate various test conditions related to time pressure, distraction, and mental workload, we introduce two extra experimental tools, timer and secondary task. A *timer* is used to create time pressure by showing a participant how

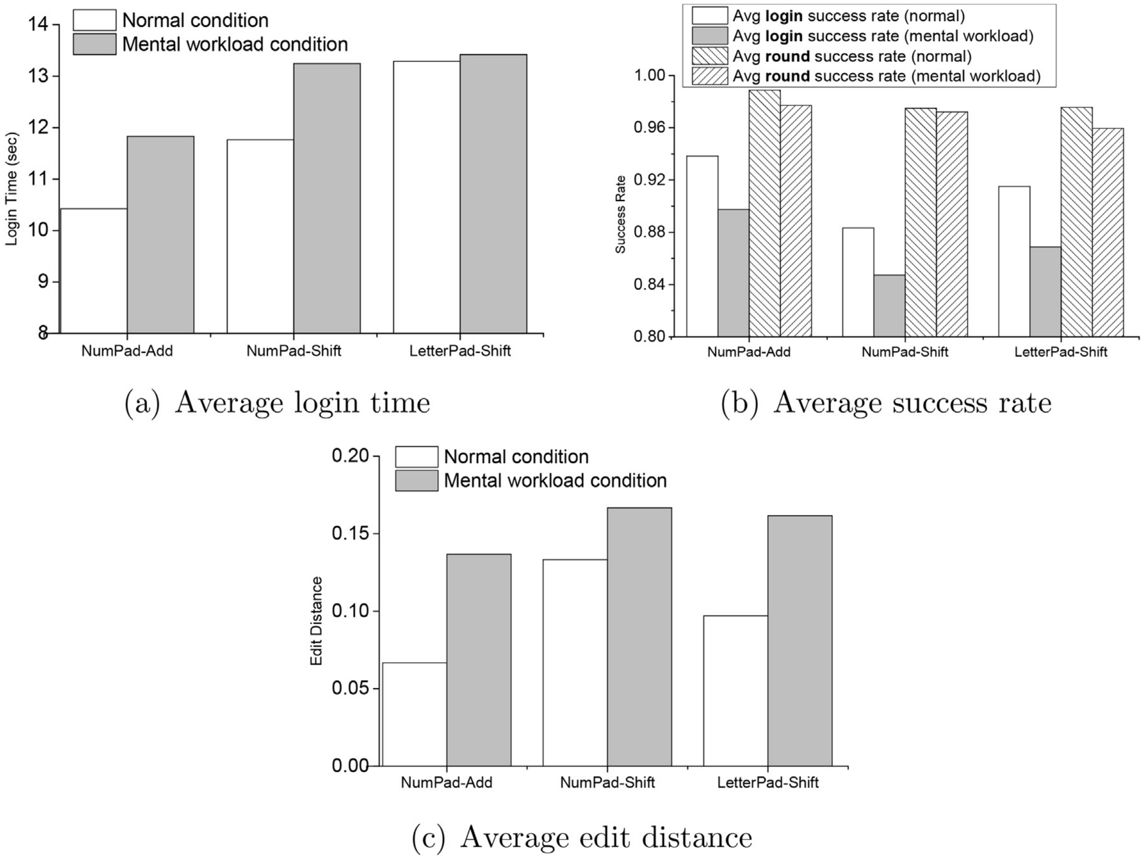


(a) Average login time

(b) Average round success rate

(c) Average edit distance

Fig. 8 – Impact of distraction.



(a) Average login time

(b) Average success rate

(c) Average edit distance

Fig. 9 – Impact of mental workload.

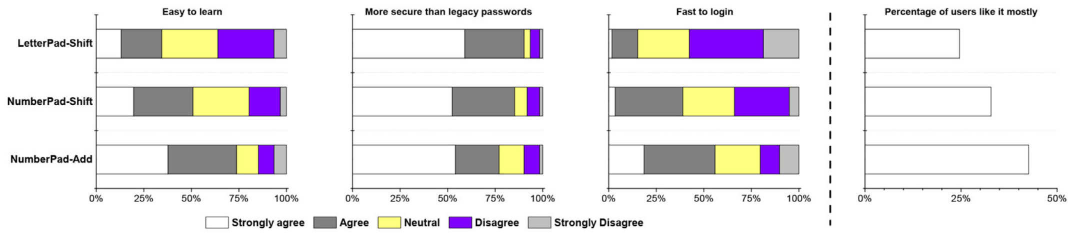


Fig. 10 – Perception of participants.

much time is left for the current test condition. It is implemented as a progress bar whose length increases every second with a countdown text field showing how many seconds are left. *Secondary tasks* are used to simulate unexpected distraction and persistent mental workload. We use *choice reaction time* (CRT) tasks as secondary tasks, which is a standard technology in experimental psychology (Jensen, 1987). CRT tasks usually work as secondary tasks that occupy the central executive[3] in human brain when evaluating the performance of a primary task in the presence of a secondary task. CRT tasks require participants to give distinct responses for each possible stimulus. In our implementation, the participants are asked to press the correct button among *N* buttons, where the correct button should have the same color as the stimulus. For example, if the stimulus shows a red button, a participant should press the red button among *N* buttons with different colors. We use $N = 2$ for tests in the distraction condition as the major focus is to unexpectedly disrupt password entry with a CRT task. We use $N = 8$ for tests in the mental workload condition so as to create a considerable mental workload, which is the same as in the classic Jensen Box setting (Jensen, 1987).

Based on the above experimental tools, we simulate six test conditions for each test group by combining two modes

[3] The central executive is a control system that mediates attention and regulation of processes occurring in working memory (Baddeley and Hitch, 1974).

Table 1 – Comparison between CoverPad and legacy passwords using usability-deployability-security metrics (Bonneau et al., 2012). • = offer the benefit, ◦ = almost offer the benefit, *no circle* = does not offer the benefit.

	Nothing-to-carry	Easy-to-learn	Efficient-to-use	Infrequent-errors	Easy-recovery-from-loss	Accessible	Negligible-cost-per-user	Server-compatible
CoverPad Schemes	•	•	◦	◦	•	•	•	
Legacy Passwords	•	•	•	◦	•	•	•	•

and three statuses. The two modes related to a timer are described as follows:

- **Relaxed mode**: A participant is asked to minimize the error rate in a fixed number of login attempts where time is not considered in performance score calculation. The number of login attempts is 5 for no-extra-task status and 3 for distraction and mental workload statuses.
- **Timed mode**: A participant is asked to perform as many successful logins as possible within 1 min, where both time and accuracy are considered in performance score calculation. The countdown of a timer creates time pressure.

Three statuses related to secondary tasks are described as follows:

- **No-extra-task status**: A participant is asked to perform the login task only.
- **Distraction status**[4]: A simple CRT task may appear with 1/3 probability each time when a participant presses a response key. This task is used to create unexpected distractions during password entry.
- **Mental workload status**: A relatively complex CRT task appears every time when a participant presses a response key. This task is used to create continuing mental workload during password entry.

Among six conditions, we referred to the combination of *relaxed* mode and *no-extra-task* status as the **normal condition**, which is the common condition usually tested in prior work (Hopper and Blum, 2001; Li and Shum, 2005; Weinshall, 2006; Wiedenbeck et al., 2006; Bai et al., 2008; Kumar et al., 2007; Sasamoto et al., 2008; De Luca et al., 2009a, b; Kim et al., 2010; Bianchi et al., 2011b, a).

7.3. Experimental results

We measure user performance with the following *metrics*: average login time, login success rates, round success rates, and average edit distances. A *round* success rate is the average success rate for a user to correctly input one password element by applying a hidden transformation. An *edit distance* is the minimum number of insertions, deletions, substitutions, and adjacent transpositions required to transform an input string into the correct password string so that an *average* edit distance is the average value of edit distances calculated from all login attempts of a user under a test condition. Among these metrics, login success rates, round success rates, and average edit distances are used to evaluate *login accuracy*.

We use the following statistical tools to test the significance of our experimental results, where a significance level of $\alpha = 0.05$ is used. For each comparison, we run an *omnibus* test across all test conditions for each scheme. Since all our performance data are quantitative, we use *Kruskal–Wallis* (KW) test for omnibus tests, which is an analogue of ANOVA but does not require normality. If the omnibus test indicates significance, we further use *Mann–Whitney* (MW) *U* test to perform pair-wise comparisons so as to identify specific pairs with significant differences. The detailed results of our statistical tests are given in Yan et al. (2013).

7.3.1. Performance under normal condition

In the normal condition, a participant is only asked to perform login tasks without any time pressure or secondary tasks. It corresponds to the combination of relaxed mode and no-extra-task status, which is used as a *baseline* in our tests.

Fig. 6(a) shows the average time for a successful login attempt in the normal condition. For all the three schemes, most participants are able to finish the login within 13 s. Although the login time of our schemes are shorter than the prior schemes (Hopper and Blum, 2001; Li and Shum, 2005; Weinshall, 2006; Wiedenbeck et al., 2006; Bai et al., 2008; Kumar et al., 2007; De Luca et al., 2009a; Kim et al., 2010; Bianchi et al., 2011b, a), it is still slower than legacy passwords. Nonetheless, it is sufficiently fast for security-sensitive applications such as online banking and corporate login, comparing to the login time of hardware-based OTPs. Fig. 6(b) and (c) show the corresponding login accuracy. Since our experiment limits the number of login attempts to 5 in order to prevent the participants from feeling exhausted or bored, even a single mistake would take the login success rate down to 80%. Our results indicate that most participants make *at most* one mistake when they use our schemes for the first time after a short training. This is shown by 97.5% average round success rate and 0.13 average edit distance in the worst case. Particularly, for the distribution of average edit distance of NumPad-Shift, 27 participants among 40 samples[5] have an average edit distance of zero (i.e. no mistakes during all tests under the test condition), which are

[4] The design of this distraction condition is different from the prior work (Dunphy et al., 2008), where the only environmental noise was simulated by playing videos and sounds that a participant does not have to respond.

[5] In order to neutralize the influence of the learning curve, we removed the experimental data when NumPad-Shift appears as the first test group. More details and discussion can be found in Yan et al. (2013).

Browser-compatible	Mature	Non-proprietary	Resilient-to-physical-observation	Resilient-to-targeted-impersonation	Resilient-to-internal-observation	Resilient-to-theft	No-trusted-third-party	Requiring-explicit-consent	Unlinkable
●		●	●	○	○	●	●	●	●
●	●	●		○		●	●	●	●

shown as a cluster of *outliers* at the bottom of the box chart. The login accuracy is expected to increase after the participants get more familiar with the schemes.

7.3.2. *Influence of time pressure*

Fig. 7 shows the impact of time pressure without any secondary tasks. The results show that the participants behave much hastily in the presence of time pressure. The average time for a successful login attempt becomes shorter and the login accuracy is decreased. The statistical tests show the difference in login time is significant ($p = 0.017$ for NumPad-Add and $p < 0.001$ for LetterPad-Shift) but the difference in login accuracy is not.

The insignificant results in login accuracy are due to the *ceiling* effect, which implies the tests are not sufficiently difficult to distinguish the influence of different test conditions. This effect could be caused by our scheme design, which is not difficult for the participants to use so that the majority of the participants did not make any mistakes during all the tests. However, even without statistical significance, we still observe the average results of login accuracy become worse with time pressure for all three tested schemes. Considering the simple design of our schemes, this indicates that time pressure may have larger influence on login accuracy if a more complex scheme is in use.

7.3.3. *Influence of distraction*

Fig. 8 shows the impact of distraction without time pressure. Many participants made a mistake when they saw a distraction task for the first time (however, NumPad-Shift is an exception). For NumPad-Add and LetterPad-Shift shown in Fig. 8(b), the round success rate returns to a comparable level as the normal condition, after the first time the distraction task appears. This indicates that the distraction task is no longer a surprise for the participants. However, even after the participants get familiar with the distraction tasks, compared to the normal condition, the success rate is still lower, the average edit distance is larger, and the average login time is longer. Nonetheless, statistical tests show that these differences are not significant.

7.3.4. *Influence of mental workload*

Fig. 9 shows the impact of mental workload without time pressure. The average login time becomes significantly longer with mental workload ($p = 0.003$ for NumPad-Add) due to context switch in users' mind between password inputs and secondary CRT tasks. An extra startup time is required to release the central executive after each CRT task. Our experiment simulates the case when users cannot get rid of other thoughts during password entry. The actual effect of mental workload depends on the status of users' mind. The impact may be elevated when the actual mental workload is higher than our CRT tasks. On the other hand, the login accuracy is lower compared to the normal condition but the difference is not significant due to the same ceiling effect mentioned in Section 7.3.2.

7.3.5. *Performance under combined conditions*

We also examine the overall impact when distraction or mental workload appears together with time pressure. As expected, compared to their counterparts without time pressure, the average login time becomes shorter (from 10.3 s to 11.7 s on average), the login success rate becomes even lower (from 87.5% to 81.3%), and the average edit distance becomes larger (from 0.151 to 0.243). Statistical tests show the difference in login time is significant ($p = 0.009$ for NumPad-Add, $p = 0.019$ for NumPad-Shift, and $p < 0.001$ for LetterPad-Shift) and the difference in login accuracy is still not significant due to the ceiling effect explained in Section 7.3.2. These results show that time pressure is still an effective stimulus to speed up password entry even in the presence of secondary tasks.

7.3.6. *User perception*

Fig. 10 shows the perception of participants collected from questionnaires. The results indicate that the participants generally feel that our schemes are secure and easy to use. While NumPad-Add is the most popular, the other two schemes also have their favorite users. Since the age of the participants are close, we further perform a group of statistical tests for the difference between males and females. The tests show that the difference is not significant for both login performance and questionnaire feedback.

7.4. ***Comparison with legacy passwords***

Table 1 gives a comparison between CoverPad and legacy passwords based on the *usability-deployability-security* metrics proposed in Bonneau et al. (2012), where a metric is not shown if neither our schemes nor legacy passwords offer corresponding benefit. Overall, this table shows that our schemes significantly improve the security strength while retaining most benefits of legacy passwords. The important benefits retained include *nothing-to-carry*, *easy-to-learn*, *easy-recovery-from-loss*, *no-trusted-third-party*, and *unlinkable* (not linked to a user's real identity). In particular, we have the following observations in comparison.

- Our schemes are rated as not *mature* since they are recently proposed and have not been widely deployed.

- Our schemes are not *server-compatible*, as most current servers support only static and replayable passwords, which may change in the near future.
- Our schemes are *quasi-resilient-to-internal-observation* in a sense that any key logger or malware which fails to capture the hidden transformation causes no password leakage.

More comparison between our schemes and prior schemes can be found in Section 2.

7.5. *Limitations*

Ecological validity is a challenging issue in any user study. Like most prior research (Hopper and Blum, 2001; Li and Shum, 2005; Kumar et al., 2007; De Luca et al., 2009b; Kim et al., 2010), our experiments engage only university students. These participants are younger and more educated compared to the general population. Therefore, usability evaluation may vary with other populations. Our experiments are also restricted by the sample size, which may affect the results of statistical tests. Although the current evaluation is only conducted on tablets, our scheme can also be used for mobile phones, as shown in Yan et al. (2013). Moreover, our user study does not include experiments on memory effects (e.g. forgetting). Since our scheme uses the same alphabet and password composition as legacy passwords, users may use the same coping strategies to help themselves to memorize the passwords in our scheme. The impact of memory effects on the user performance would be similar to legacy passwords as shown in the prior literature (De Luca et al., 2010).

8. Conclusions

In this paper, we investigated the underlying challenges of designing LRPE schemes, and developed a broad set of design criteria for secure and usable LRPE schemes, which cover security–usability relations, built-in security, and universal accessibility. Guided by these criteria, we proposed a practical LRPE scheme leveraging on the touchscreen feature of mobile devices. It improves leakage resilience while preserving most benefits of legacy passwords. The practicability of our scheme was verified in an extended user study that incorporates new experiments to examine the influence of additional test conditions, where time pressure and mental workload were shown to have significant impacts on user performance. We expect this new design of user experiments along with the proposed design criteria provide insights into the future development of LRPE schemes.

Acknowledgments

The study of Yingjiu Li on this subject is funded through a research grant C220/MSS12C003 from MOE's AcRF Tier 1 funding support through Singapore Management University.

REFERENCES

Androidcommunity. Samsung galaxy siii display specs. 2012. online [accessed 13.05.14], http://androidcommunity.com/samsung-galaxy-siii-display-specs-edge-out-iphone-5-20121002/.

Asghar HJ, Li S, Pieprzyk J, Wang H. Cryptanalysis of the convex hull click human identification protocol. In: Proceedings of the 13th International Conference on Information Security; 2010. p. 24–30.

Asghar HJ, Li S, Steinfeld R, Pieprzyk J. Does counting still count? revisiting the security of counting based user authentication protocols against statistical attacks. In: Proceedings of the 20th Annual Network and Distributed System Security Symposium; 2013.

Aumann RJ. Subjectivity and correlation in randomized strategies. J Math Econ 1974;1(1):67–96.

Baddeley AD, Hitch G. Working memory. Psychol Learn Motiv 1974;8:47–89.

Bai X, Gu W, Chellappan S, Wang X, Xuan D, Ma B. Pas: predicate-based authentication services against powerful passive adversaries. In: Proceedings of the 2008 Annual Computer Security Applications Conference; 2008. p. 433–42.

Begemann O. Remote view controllers in iOS 6. 2012. online [accessed 13.05.2014], http://oleb.net/blog/2012/10/remote-view-controllers-in-ios-6.

Bianchi A, Oakley I, Kostakos V, Kwon DS. The phone lock: audio and haptic shoulder-surfing resistant pin entry methods for mobile devices. In: Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction; 2011. p. 197–200.

Bianchi A, Oakley I, Kwon D-S. Obfuscating authentication through haptics, sound and light. In: Proceedings of the 2011 Annual Conference extended abstracts on Human factors in Computing Systems; 2011. p. 1105–10.

Biddle R, Chiasson S, van Oorschot PC. Graphical passwords: learning from the first twelve years. ACM Comput Surv 2012;44(4).

Bonneau J, Herley C, van Oorschot P, Stajano F. The quest to replace passwords: a framework for comparative evaluation of web authentication schemes. In: Proceedings of IEEE Symposium on Security and Privacy; 2012.

Bright P. RSA finally comes clean: SecurID is compromised. 2011. online [accessed 13.05.14], http://arstechnica.com/security/news/2011/06/rsa-finally-comes -clean-securid-is-compromised.ars.

Colavita FB. Human sensory dominance. Atten Percept Psychophys 1974;16(2):409–12.

Coskun B, Herley C. Can "something you know" be saved?. In: Proceedings of the 11th International Conference on Information Security; 2008. p. 421–40.

De Luca A, Denzel M, Hussmann H. Look into my eyes!: can you guess my password?. In: Proceedings of the 5th Symposium on Usable Privacy and Security; 2009. 7:1–7:12.

De Luca A, Langheinrich M, Hussmann H. Towards understanding atm security: a field study of real world atm use. In: Proceedings of the Sixth Symposium on Usable Privacy and Security; 2010.

De Luca A, von Zezschwitz E, Husmann H. Vibrapass: secure authentication based on shared lies. In: Proceedings of the 27th International Conference on Human factors in computing systems; 2009. p. 913–6.

Dunphy P, Fitch A, Olivier P. Gaze-contingent passwords at the atm. In: Proceedings of the 4th Conference on Communication by Gaze Interaction; 2008.

Gibson JJ. Adaptation, after-effect, and contrast in the perception of curved lines. J Exp Psychol 1937;20(6):553–69.

Ginzburg, L., Sitar, P., Flanagin, G. K. (2010). User authentication system and method. US Patent 7,725,712, SyferLock Technology Corporation.

Golle P, Wagner D. Cryptanalysis of a cognitive authentication scheme (extended abstract). In: Proceedings of the 2007 IEEE Symposium on Security and Privacy; 2007. p. 66–70.

Google. Google glass explorer program. 2014. online [accessed 13.05.14], http://www.google.com/glass/start/how-to-get-one.

Hopper NJ, Blum M. Secure human identification protocols. In: Proceedings of the 7th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology; 2001. p. 52–66.

Jensen AR. Process differences and individual differences in some cognitive tasks. Intelligence 1987;11(2):107–36.

Kim D, Dunphy P, Briggs P, Hook J, Nicholson JW, Nicholson J, et al. Multi-touch authentication on tabletops. In: Proceedings of the 28th International Conference on Human factors in Computing Systems; 2010. p. 1093–102.

Kumar M, Garfinkel T, Boneh D, Winograd T. Reducing shoulder-surfing by using gaze-based password entry. In: Proceedings of the 3rd Symposium on Usable Privacy and Security; 2007. p. 13–9.

Li S, Asghar H, Pieprzyk J, Sadeghi A-R, Schmitz R, Wang H. On the security of pas (predicate-based authentication service). In: Proceedings of the 2009 Annual Computer Security Applications Conference; 2009. p. 209–18.

Li S, Shum H-Y. Secure human-computer identification (interface) systems against peeping attacks: SecHCI. In: Cryptology ePrint Archive, Report 2005/268; 2005.

Long J, Wiles J. No Tech Hacking: a guide to social engineering, dumpster diving, and shoulder surfing. Syngress; 2008.

Matsumoto T. Gummy and conductive silicone rubber fingers. In: Proceedings of the 8th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology; 2002. p. 574–6.

Matsumoto T, Imai H. Human identification through insecure channel. In: Proceedings of the 10th Annual International Conference on Theory and Application of Cryptographic Techniques; 1991. p. 409–21.

Microsoft. Windows 8. 2014. online [accessed 13.05.14], http://windows.microsoft.com.

Perkovic T, Mumtaz A, Javed Y, Li S, Khayam SA, Cagalj M. Breaking undercover: exploiting design flaws and nonuniform human behavior. In: Proceedings of the Seventh Symposium on Usable Privacy and Security; 2011.

Roth V, Richter K, Freidinger R. A pin-entry method resilient against shoulder surfing. In: Proceedings of the 11th ACM conference on Computer and Communications Security; 2004. p. 236–45.

RSA. RSA SecurID two-factor authentication. 2011. online [accessed 13.05.14], http://www.rsa.com/products/securid/sb/10695_SIDTF A_SB_0210.pdf.

Sasamoto H, Christin N, Hayashi E. Undercover: authentication usable in front of prying eyes. In: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human factors in Computing Systems; 2008. p. 183–92.

Schenker L. Pushbutton calling with a two-group voice-frequency code. Bell Syst Tech J 1960;39(1):235–55.

Song DX, Wagner D, Tian X. Timing analysis of keystrokes and timing attacks on ssh. In: Proceedings of the 10th USENIX Security Symposium; 2001.

Stifelman LJ, Arons B, Schmandt C, Hulteen EA. Voicenotes: a speech interface for a hand-held voice notetaker. In: Proceedings of the INTERCHI '93 conference on Human factors in computing systems; 1993. p. 179–86.

Weinshall D. Cognitive authentication schemes safe against spyware (short paper). In: Proceedings of the 2006 IEEE Symposium on Security and Privacy; 2006. p. 295–300.

Wiedenbeck S, Waters J, Sobrado L, Birget J-C. Design and evaluation of a shoulder-surfing resistant graphical password scheme. In: Proceedings of the Working Conference on Advanced Visual Interfaces; 2006. p. 177–84.

Yan Q, Han J, Li Y, Deng RH. On limitations of designing leakage-resilient password systems: attacks, principles and usability. In: Proceedings of the 19th Annual Network and Distributed System Security Symposium; 2012.

Yan Q, Han J, Li Y, Zhou J, Deng RH. Designing leakage-resilient password entry on touchscreen mobile devices. In: Proceedings of the 8th ACM Symposium on Information, Computer and Communications Security; 2013. p. 37–48.

Qiang Yan is currently a Software Engineer at Google Inc. He received Ph.D. in Information Systems from Singapore Management University. His research focuses on system and network security in general. He received both his M.Sc. in Computer Science and B.Eng. in Software Engineering from Fudan University, P.R. China.

Jin Han is a Research Scientist at Institute for Infocomm Research, Singapore. Jin's current research interests mainly focus on mobile security and leakage-resilient password systems. Jin received his Ph.D. in Information Systems from Singapore Management University. He received both his M.Sc. in Computer Science and B.Eng. in Software Engineering from Fudan University, P.R. China.

Yingjiu Li is currently an Associate Professor in the School of Information Systems at Singapore Management University. His research interests include RFID Security and Privacy, Mobile and System Security, Applied Cryptography and Cloud Security, and Data Application Security and Privacy. Yingjiu Li is a senior member of the ACM and a member of the IEEE Computer Society. The URL for his web page is http://www.mysmu.edu/faculty/yjli

Jianying Zhou is the head of Infocomm Security Department at Institute for Infocomm Research. He received Ph.D. in Information Security from University of London. His research interests are in computer and network security, mobile and wireless security. He is a founder and steering committee member of International Conference on Applied Cryptography and Network Security (ACNS).

Robert H. Deng has been a Professor at the School of Information Systems, Singapore Management University since 2004. Prior to this, he was a Principal Scientist at the Institute for Infocomm Research, Singapore. His research interests include data security and privacy, multimedia security, network and system security. He received the University Outstanding Researcher Award from the National University of Singapore in 1999 and the Lee Kuan Yew Fellow for Research Excellence from the Singapore Management University in 2006. He was named Community Service Star and Showcased Senior Information Security Professional by (ISC)^2 under its Asia–Pacific Information Security Leadership Achievements program in 2010.